



# A proof for loop-law constraints in stoichiometric metabolic networks

## Citation

Noor, Elad, Nathan E Lewis, and Ron Milo. 2012. A proof for loop-law constraints in stoichiometric metabolic networks. BMC Systems Biology 6: 140.

## Published Version

doi:10.1186/1752-0509-6-140

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11180365>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

RESEARCH ARTICLE

Open Access

# A proof for loop-law constraints in stoichiometric metabolic networks

Elad Noor<sup>1\*</sup>, Nathan E Lewis<sup>2,3\*</sup> and Ron Milo<sup>1</sup>

## Abstract

**Background:** Constraint-based modeling is increasingly employed for metabolic network analysis. Its underlying assumption is that natural metabolic phenotypes can be predicted by adding physicochemical constraints to remove unrealistic metabolic flux solutions. The loopless-COBRA approach provides an additional constraint that eliminates thermodynamically infeasible internal cycles (or loops) from the space of solutions. This allows the prediction of flux solutions that are more consistent with experimental data. However, it is not clear if this approach over-constrains the models by removing non-loop solutions as well.

**Results:** Here we apply Gordan's theorem from linear algebra to prove for the first time that the constraints added in loopless-COBRA do not over-constrain the problem beyond the elimination of the loops themselves.

**Conclusions:** The loopless-COBRA constraints can be reliably applied. Furthermore, this proof may be adapted to evaluate the theoretical soundness for other methods in constraint-based modeling.

**Keywords:** Thermodynamics, Gordan's theorem, Metabolism

## Background

Constraint-based modeling has become a successful framework for the analysis of large and complex stoichiometric biochemical networks [1]. The underlying concept of this framework is that one can use the stoichiometry of each reaction in a reconstructed metabolic network [2,3] and known bounds on reaction fluxes to compute metabolic flux for each reaction. These predictions represent allowable steady-state metabolic flux distributions in a cell under a given growth condition. Some example constraints include mass balance and metabolite uptake rates. One set of constraints that has been more challenging to implement are those associated with thermodynamic limitations. Without thermodynamic constraints, non-physical fluxes can be computed for some metabolic reactions if they produce an internal cycle (Figure 1a). Such cycles of reactions violate a "loop law" that is analogous

to Kirchhoff's second law for electrical circuits, as discussed previously by Beard et al. [4]. Many approaches have successfully constrained these loops using known flux directionality [5], energy-balance equations [4], and known [6-11] or predicted [12] thermodynamic parameters. Loops have also been indirectly removed by minimizing network flux [6,13-15], or by coupling flux to enzyme synthesis costs [16].

A new approach, called loopless-COBRA, was recently presented [17]. Unlike previous loop-removal algorithms, this method does not necessitate extra inputs or data, such as metabolite concentrations or thermodynamic parameters. Basically, this method imposes the second law of thermodynamics by using a mixed-integer linear programming (MILP) approach to constrain flux solutions so that they obey the loop law. Thus, flux solutions from this method are all within the portion of the flux space that is devoid of loops (Figure 1b). Excitingly, the loop removal improved the consistency of the simulation results [17] with experimental data [18]. Specifically, it provides more realistic flux values for reactions that normally contribute to loops in a model. Otherwise, flux predictions for such reactions would usually have to be ignored in any subsequent analysis.

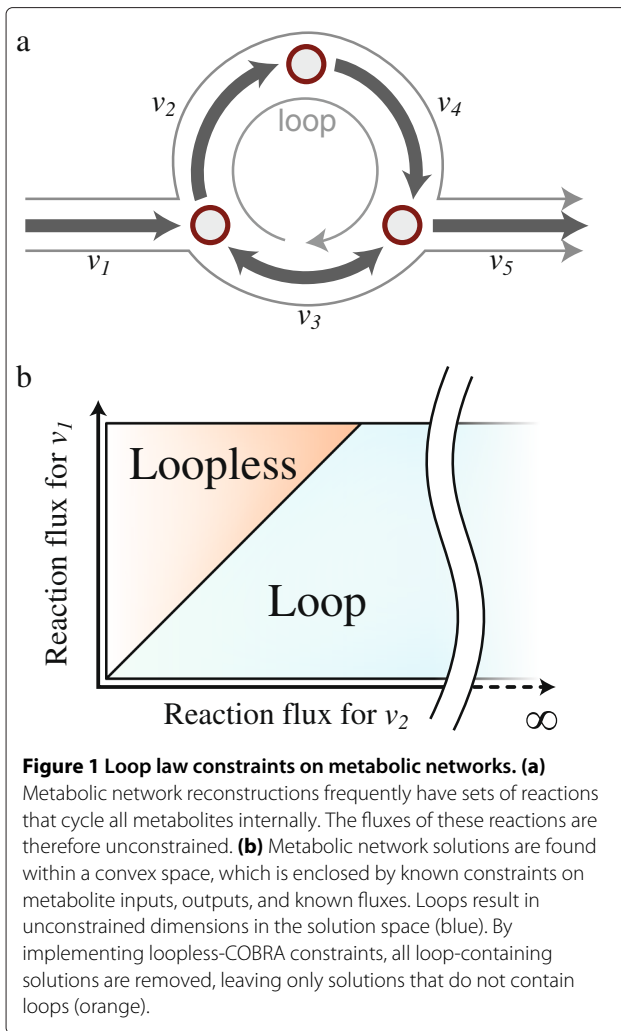
\*Correspondence: elad.noor@weizmann.ac.il;

nlewis@genetics.med.harvard.edu

<sup>1</sup> Department of Plant Sciences, 234 Herzl st., Weizmann Institute of Science, Rehovot 76100, Israel

<sup>2</sup> Department of Bioengineering, 9500 Gilman Drive, University of California San Diego, La Jolla, CA 92093-0412, USA

Full list of author information is available at the end of the article



The loopless-COBRA method was shown to work in various scenarios, and the paper that presented the approach provides an explanation for why this method works. However, there is no mathematical proof for its formulation as an optimization problem. Specifically, it does not demonstrate that the additional MILP constraints do not over-constrain the problem and eliminate some non-loop containing solutions. Since constraint-based methods attempt to only eliminate impossible *in silico* phenotypes (i.e., steady-state flux distributions that the cell cannot maintain), it is important to verify that solutions representing real phenotypes are not removed by accidentally over-constraining the problem.

Here, we address this issue by presenting a mathematical proof for the completeness and soundness of the loopless-COBRA method, thereby adding fundamental support and rigorous proof for the constraints presented by Schellenberger et al. [17].

## Results and discussion

### Formal definition of loop-law constraints

In loopless-COBRA, the constraints added to the linear problem are:

$$G_i < 0 \quad \text{for all} \quad v_i > 0$$

$$G_i > 0 \quad \text{for all} \quad v_i < 0$$

$$G_i \in \mathbb{R} \quad \text{for all} \quad v_i = 0$$

$$N_{int}G = 0$$

where  $v_i$  are the flux variables and  $N_{int}$  is the null-space matrix of  $S_{int}$  (the stoichiometric matrix of internal reactions). The third constraint ( $G_i \in \mathbb{R}$ ) is not actually a constraint, but a way to say that  $G_i$  can have any value if  $v_i = 0$ . We can rewrite all these constraints succinctly as:

$$\exists G \in \mathbb{R}^n \text{ s.t. } \text{null}(S_{int}) \cdot G = 0 \wedge (\forall i \text{ sign}(G_i) = -\text{sign}(v_i) \vee v_i = 0) \quad (1)$$

### Formal definition of loops

In order to prove that this constraint eliminates loops (and only loops), we must first find a mathematical formulation for a loop, using the same notation as above. We thus define a loop as a nonzero vector  $x \in \mathbb{R}^n$  which satisfies the mass-balance equation for the internal reactions, i.e.  $S_{int} \cdot x = 0$ . This means that although there is a nonzero net flux in some of the reactions, overall, the internal network is at steady-state (an obvious violation of the second law of thermodynamics). It is important to point out, that this equation for defining loops must not be confused with the steady-state assumption commonly used in flux balance analysis models, namely  $S \cdot v = 0$ , where the *full* stoichiometric matrix ( $S$ ) is used.

According to this definition, a flux distribution ( $v \in \mathbb{R}^n$ ) will contain a loop if and only if there exists a vector  $x \in \mathbb{R}^n \setminus \{0\}$  which is consistent with the flux directions in  $v$  (i.e.  $x_i$  is either zero or has the same sign as  $v_i$ ) and is itself a loop (i.e.  $S_{int} \cdot x = 0$ ). Formally,  $v$  has a loop if and only if:

$$\exists x \in \mathbb{R}^n \setminus \{0\} \text{ s.t. } (\forall i \text{ sign}(x_i) \in \{\text{sign}(v_i), 0\}) \wedge S_{int} \cdot x = 0 \quad (2)$$

We have now finished laying the groundwork for our mathematical proof that loopless-COBRA is sound and complete. In order to do that, we are left only to show that Equation 1 is satisfiable if and only if Equation 2 is unsatisfiable (in other words, there are no loops).

### Gordan's theorem

We start our proof by quoting Gordan's theorem: For all  $A \in \mathbb{R}^{m \times n}$  exactly one of the following two statements is true:

$$(a) \quad \exists x \in \mathbb{R}_+^n \setminus \{0\} \text{ s.t. } Ax = 0$$

$$(b) \quad \exists y \in \mathbb{R}^m \text{ s.t. } A^T y > 0$$

We will show that statement (a) in Gordan's theorem is equivalent to having a loop (Equation 2) and statement (b) is equivalent to the MILP constraints used by loopless-COBRA (Equation 1). After doing so, we would easily reach the conclusion of the proof.

As a guidance for the following sections, one can see that statement (a) already resembles Equation 2 if we define  $A = S_{int}$ . The only difference is that  $x$  is constrained to have only non-negative values (note the '+' in  $\mathbb{R}_+^n$ ), instead of being consistent with the sign of  $v$ . Corollaries 1 and 2 will show how we can overcome this discrepancy by defining  $A$  in a slightly different way.

At first glance, statement (b) might look unrelated to Equation 1. However, in the last part of our proof, we show that choosing  $G \in \mathbb{R}^n$ , which satisfies  $\text{null}(S_{int}) G = 0$ , is the same as choosing  $y \in \mathbb{R}^m$  and then taking  $G = S_{int}^\top \cdot y$ . Only for sake of understanding the algebra, one can think of  $y$  as the vector of formation Gibbs energies, and of  $G$  as the vector of reaction Gibbs energies. The rest of the proof, like for statement (a), deals with adjusting  $A$  to fit with the non-positive values in  $v$ . The toy example in Figure 2 shows how Gordan's theorem corresponds to having or not having a loop, for a network with 3 compounds and 3 internal reactions.

### Corollary 1

For all  $A \in \mathbb{R}^{m \times n}$  and  $d \in \{-1, 0, 1\}^n$  exactly one of the following two statements is true:

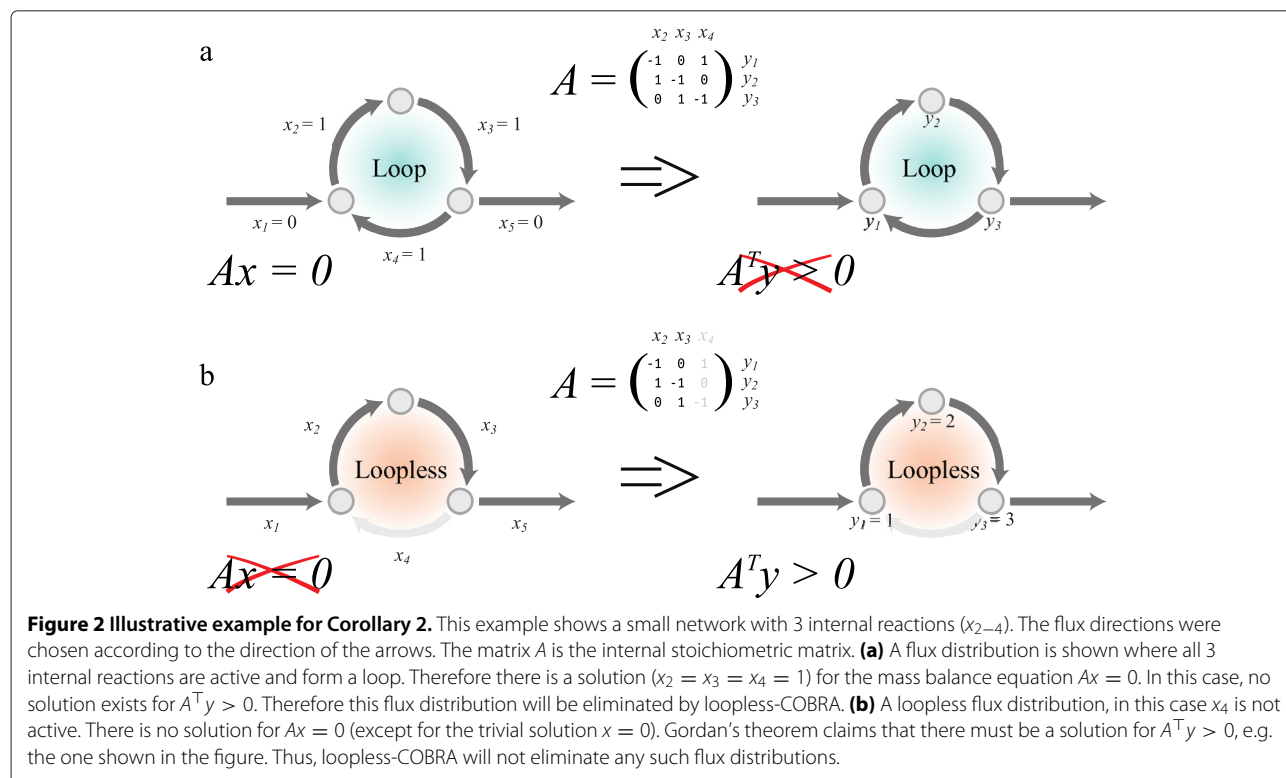
$$(1a) \quad \exists x \in \mathbb{R}^n \setminus \{0\} \quad \text{s.t.} \quad (\forall i \text{ sign}(x_i) \in \{d_i, 0\}) \wedge Ax = 0$$

$$(1b) \quad \exists y \in \mathbb{R}^m \quad \text{s.t.} \quad (\forall i \text{ sign}(A^\top y)_i = d_i \vee d_i = 0)$$

*Proof.* First, define a new matrix  $\hat{A}$  that is the same as  $A$ , without the columns corresponding to  $d_i = 0$  and where columns corresponding to  $d_i = -1$  are multiplied by  $-1$ . Statement (1a) is true for  $A$  if and only if (a) is true for  $\hat{A}$ . The forward direction is easily shown by removing the zeros from  $x$ , where  $d_i = 0$ , and negating values corresponding to  $d_i = -1$  (as previously done for  $A$ ). Reversing this process (i.e. taking a positive solution for  $\hat{A}\hat{x} = 0$ , adding back the zeros and negating the same values) shows the other direction is true as well.

Likewise, statement (1b) for  $A$  is true if and only if (b) is true for  $\hat{A}$ , since columns with  $d_i = -1$  are negated in  $\hat{A}$  and thus  $\text{sign}(\hat{A}^\top y)_i = 1$ . Columns with  $d_i = 0$  have no other constraints in (1b) and the same goes for (b) since they are removed from  $\hat{A}$ .

Therefore, Corollary 1 is directly derived from Gordan's theorem.  $\square$



Since constraint-based models usually use a vector of real values ( $v \in \mathbb{R}^n$ ) to represent the flux distribution, we subsequently change the formulation of the Corollary slightly to match.

### Corollary 2

For all  $A \in \mathbb{R}^{m \times n}$  and  $v \in \mathbb{R}^n$ , exactly one of the following two statements is true:

$$(2a) \quad \exists x \in \mathbb{R}^n \setminus \{0\} \quad \text{s.t.} \quad (\forall i \text{ sign}(x_i) \in \{\text{sign}(v_i), 0\}) \wedge Ax = 0$$

$$(2b) \quad \exists y \in \mathbb{R}^m \quad \text{s.t.} \quad (\forall i \text{ sign}(A^T y)_i = -\text{sign}(v_i) \vee v_i = 0)$$

*Proof.* Defining  $d_i \equiv \text{sign}(v_i)$ , we get this directly from Corollary 1. Note that  $-\text{sign}(v_i)$  can be used in (2b), since the existence of  $y$  is equivalent to the existence of  $-y$ .  $\square$

This adjustment now allows us to apply Corollary 2 to constraint-based problems and show that it eliminates loops (see example in Figure 2). In order to avoid any confusion, we point out that a solution for  $Ax = 0$  is considered a loop only if  $A$  is the stoichiometric matrix of *internal* reactions. This should not be mistaken as the steady-state mass-balance equation which looks exactly the same, except that  $A$  contains both internal and external reactions.

### Corollary 3

Adding the following constraint:

$$\exists y \in \mathbb{R}^m \quad \text{s.t.} \quad (\forall i \text{ sign}(A^T y)_i = -\text{sign}(v_i) \vee v_i = 0)$$

is equivalent to eliminating all loops in a flux distribution  $v$ .

*Proof.* Using Corollary 2, all that must be shown is that statement (2a) is equivalent to having a loop. This is apparent since  $x$  is a vector in the null-space of  $A$  (i.e., a loop) and is consistent with the flux direction of  $v$  in each of its nonzero reactions.  $\square$

Note that the trivial case  $v = 0$  can still be a solution and it should be explicitly avoided if necessary.

### Applying Corollary 3 in loopless-COBRA

The added constraints in loopless-COBRA [17] are slightly different than in Corollary 3, namely:

$$\exists G \in \mathbb{R}^n \quad \text{s.t.} \quad \text{null}(A) \cdot G = 0 \wedge (\forall i \text{ sign}(G_i) = -\text{sign}(v_i) \vee v_i = 0)$$

We claim here that both formulations are equivalent. The fundamental theorem of linear algebra states that the nullspace,  $\text{null}(A)$ , is the orthogonal complement of the row space,  $\text{image}(A^T)$ . Therefore, we can say that  $\text{null}(A) \cdot$

$G = 0$  if and only if  $G \in \text{image}(A^T)$ , so we can rewrite the constraint above as:

$$\exists G \in \text{image}(A^T) \quad \text{s.t.} \quad (\forall i \text{ sign}(G_i) = -\text{sign}(v_i) \vee v_i = 0)$$

which is obviously equivalent to the constraint in Corollary 3.

## Conclusions

Our results prove that the constraints proposed by Schellenberger, et al. [17] eliminate all flux solutions with loops and nothing more. This alleviates the concern as to if the loopless-COBRA constraints might eliminate true flux states. Furthermore, values in  $G$  are analogous to the change in Gibbs energy ( $\Delta_r G$ ) of the reactions [17], and  $y$  values are analogous to the chemical potentials (or formation energies,  $\Delta_f G$ ) of the compounds themselves. Since in most cases, there are fewer compounds than reactions ( $m < n$ ), we believe that it is convenient and intuitive to use the new formulation.

In conclusion, this proof provides theoretical credibility for the loopless-COBRA constraint. However, as with any algorithmic MILP implementation, care must still be taken with respect to numerical limitations and the convergence of the optimization algorithm.

Lastly, we believe this proof may be extended to similar methods addressing loop elimination. We also hope that similar proofs will appear for other methods, since more rigorous mathematical treatments are needed in many published algorithms in computational biology to prove or disprove their correctness.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

EN conceived of the study, participated in its design and coordination, and helped to draft the manuscript. NL participated in the design of the study and drafted the manuscript. RM participated in its design, coordination, and drafting the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We thank Or Sheffet, Bernhard Ø. Palsson and Uri Barenholz for mathematical assistance and helpful discussions. EN is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

Received: 5 July 2012 Accepted: 30 October 2012

Published: 12 November 2012

### References

- Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.** *Nat Rev Microbiol* 2012, **10**(4):291–305.
- Thorleifsson SG, Thiele I: **rBioNet: A COBRA toolbox extension for reconstructing high-quality biochemical networks.** *Bioinf (Oxford, England)* 2011, **27**(14):2009–10.
- Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat Protoc* 2010, **5**:93–121.
- Beard Da, Liang Sd, Qian H: **Energy balance for analysis of complex metabolic networks.** *Biophys J* 2002, **83**:79–86.
- Varma A, Palsson BO: **Metabolic capabilities of escherichia coli: I. synthesis of biosynthetic precursors and cofactors.** *J Theor Biol* 1993, **165**(4):477–502.

6. Holzhütter HG: **The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks.** *Eur J Biochem / FEBS* 2004, **271**(14):2905–22.
7. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Thermodynamics-based metabolic flux analysis.** *Biophys J* 2007, **92**(5):1792–1805.
8. Fleming RMT, Thiele I, Provan G, Nasheuer HP: **Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism.** *J Theor Biol* 2010, **264**(3):683–692.
9. Fleming RMT, Thiele I: **von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models.** *Bioinf (Oxford, England)* 2011, **27**(1):142–143.
10. Cogne G, Rüger M, Bockmayr A, Titica M, Dussap CG, Cornet JF, Legrand J: **A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae *Chlamydomonas reinhardtii*.** *Biotechnol Prog* 2011, **27**(3):631–40.
11. Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, Milo R: **An integrated open framework for thermodynamics of reactions that combines accuracy and coverage.** *Bioinf (Oxford, England)* 2012, **27**(15):2037–2044.
12. De Martino D, Figliuzzi M, De Martino A, Marinari E: **A scalable algorithm to explore the Gibbs energy landscape of genome-scale metabolic networks.** *PLoS Computat Biol* 2012, **8**(6):e1002562.
13. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BO: **Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models.** *Mol Syst Biol* 2010, **6**(390):1–13.
14. Klitgord N, Segre D: **The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles.** *Genome Inform Ser* 2010, **2009**(22):41–55.
15. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U: **Multidimensional optimality of microbial metabolism.** *Science* 2012, **336**(6081):601–604.
16. Thiele I, Fleming RMT, Bordbar A, Schellenberger J, Palsson BO: **Functional characterization of alternate optimal solutions of *Escherichia coli*'s transcriptional and translational machinery.** *Biophys J* 2010, **98**(10):2072–81.
17. Schellenberger J, Lewis NE, Palsson BO: **Elimination of thermodynamically infeasible loops in steady-state metabolic models.** *Biophys J* 2011, **100**(3):544–53.
18. Lewis NE, Cho BK, Knight EM, Palsson BO: **Gene expression profiling and the use of genome-scale in silico models of *Escherichia coli* for analysis: providing context for content.** *J bacteriol* 2009, **191**(11):3437–44.

doi:10.1186/1752-0509-6-140

**Cite this article as:** Noor et al.: A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Systems Biology* 2012 **6**:140.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

